Détermination de la dimension des réseaux d'essais

E. Gozé

Biométricien, division biométrie informatique IRCT-CIRAD, BP 5035, 34032 Montpellier Cedex 01, France.

Résumé

La dimension d'un essai se calcule d'après un objectif de puissance. Or, la puissance d'un essai multilocal est rarement connue, aussi bien pour la détection des interactions traitements-lieux que pour celle des effets principaux. Après un rappel de la finalité des essais multilocaux, du modèle et de la méthode d'analyse, sont données les formules de calcul de la puissance des réseaux d'essais en blocs complets. Le modèle est à effets lieux aléatoires et à effets traitements fixes. Dans le cas d'essais à 3 objets, les courbes de puissance des différents tests sont tracées et commentées . Un exemple d'optimisation économique est ensuite présenté. Pour la détection des effets principaux sur le rendement coton-graine, un gain de puissance substantiel peut être obtenu en remplaçant les blocs dispersés par des dispositifs à 2 ou 3 répétitions par lieu.

MOTS-CLES : coton, essais multilocaux, essais en milieu paysan, puissance, analyse de variance, régionalisation, interaction, optimisation.

Introduction

But

Le but d'une expérimentation multilocale est d'éprouver une ou plusieurs innovations (variétales, culturales, phytosanitaires) dans diverses conditions de milieu (climat, sol, paysan). A l'issue de cette épreuve, l'innovation peut être acceptée ou rejetée en fonction de critères qualitatifs décisifs : port d'une variété, sensibilité aux carences, toxicité d'un produit phytosanitaire, acceptabilité par l'agriculteur… Le choix de l'un ou l'autre des objets testés peut être aussi déterminé par des critères quantitatifs, tels que mesure des infestations, qualité de la fibre et rendement, en ce qui concerne la culture cotonnière. Ce sont ces critères quantitatifs qui nous intéressent ici.

L'innovation, seule ou en concurrence avec d'autres, est le plus souvent comparée à un témoin. On cherche alors à savoir si elle apporte un gain économique sur l'ensemble de la zone couverte par le dispositif multilocal (ce qui ne veut pas dire que le bilan de l'innovation soit positif pour chacun des cultivateurs de la zone). Pour être adoptée, il faut que l'innovation :

- soit apporte un gain dans toutes les situations possibles (cas A) ;

- soit apporte des gains suffisants dans la plupart des situations pour que, sur l'ensemble du territoire couvert, d'éventuelles pertes soient plus que compensées (cas B). On ne cherche pas à tirer parti des particularités locales.

Modèle

Pour pouvoir faire de l'inférence sur l'ensemble du territoire étudié, il faut conduire une expérimentation sur un échantillon de sites représentatif des situations possibles, sur lesquels on va mesurer une variable quantitative y. En l'absence de connaissances préalables mieux adaptées au phénomène étudié, on représente la réalité par un modèle linéaire assez général :

$$y_{ijk} = m + a_i + B_j + (aB)_{ij} + c_{jk} + E_{ijk}$$

m	est l'effet moyen		
a_i	est l'effet de l'objet i	(fixe)	i = 1..p objets
B_j	est l'effet du lieu j	(aléatoire)	j = 1..q lieux
$(aB)_{ij}$	est l'interaction objet x lieu	(aléatoire)	
c_{jk}	est l'effet du bloc k du lieu j	(fixe)	k=1..r blocs/lieu
E_{ijk}	est l'erreur résiduelle	(aléatoire)	

On suppose que les erreurs résiduelles suivent une loi normale d'espérance nulle, qu'elles sont indépendantes et partout de même variance, et qu'il en est de même pour les termes d'interaction. De plus, les erreurs et les interactions sont supposées indépendantes entre elles.

Analyse de variance

Les buts qu'on a décrits plus haut se ramènent à la détection d'effets principaux et d'interactions, que l'on réalise par des tests statistiques. L'analyse de variance nous permet de calculer des carrés moyens dont les espérances et les degrés de liberté sont donnés au tableau 1.

TABLEAU 1
Espérances des carrés moyens de l'analyse de variance d'un essai multilocal.
Mean square (CM) expectations for the analysis of variance for a multisite experiment.

Carré moyen	Espérance	D.d.l.
CM_a	$\sigma^2_E + r\sigma^2_{aB} + rq \sum_{i=1}^{P} \frac{a_i^2}{(p-1)}$	(p - 1)
CM_B	$\sigma^2_E + rp\sigma^2_B$	(q - 1)
CM_{aB}	$\sigma^2_E + r\sigma^2_{aB}$	(p - 1) (q - 1)
CM_E	σ^2_E	q (p - 1) (r - 1)

De même que σ^2_E est la variance des erreurs E_{ijk}, σ^2_{aB} est la variance des termes d'interaction $(aB)_{ij}$. Ces termes sont supposés suivre une loi normale.

Quand cette variance est nulle, les effets des objets et des lieux sont strictement additifs : les différences entre objets sont les mêmes dans chaque lieu, aux erreurs résiduelles près.

Quand la variance d'interaction vaut 100, en moyenne 50 % de ces $(aB)_{ij}$ dépassent en valeur absolue 0,675 x $\sqrt{100}$ = 6,75 et 5 % dépassent 1,960 x $\sqrt{100}$ = 19,6.

Remarquons que le carré moyen de l'interaction n'est pas égal à la variance d'interaction. Si on veut estimer cette variance à partir des résultats d'un essai, on utilise l'estimateur sans biais :

$$\hat{\sigma}^2_{aB} = \frac{CM_{aB} - CM_E}{r}$$

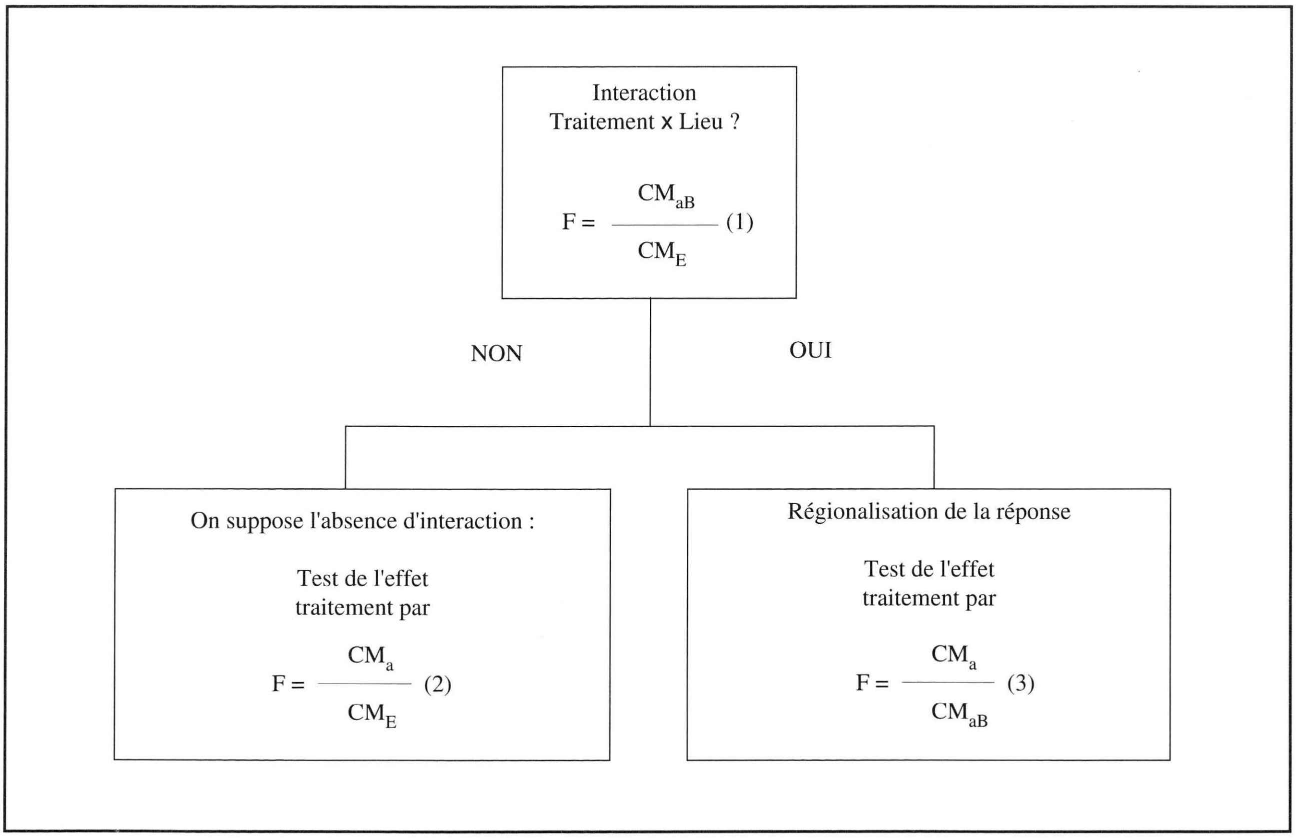


Figure 1
Conduite des tests, démarche classique.
Conventional test procedure.

Conduite des tests

Comme indiqué figure 1, quand aucune interaction n'est détectée, on en conclut habituellement l'absence d'interaction. Le gain éventuel apporté par l'innovation est alors supposé le même pour tous les lieux. Le test de l'effet traitement se fait en formant le rapport CM_a/CM_E.

En revanche, en présence d'interaction, on peut détecter un effet moyen sur l'ensemble de la zone concernée par l'expérimentation en formant le rapport CM_a/CM_{aB} : c'est ce que l'on appelle la régionalisation de la réponse.

Le problème qu'on se propose d'étudier ici est de déterminer le nombre d'essais et le nombre de blocs nécessaires pour maximiser les chances de détection de l'interaction et des effets principaux, avec le minimum de moyens.

Dans un test statistique, la probabilité de rejeter l'hypothèse nulle (ici de détecter un effet) est appelée puissance (voir à ce sujet PHILIPPEAU, 1984 ou DAGNELIE, 1973). Elle dépend de l'importance de l'écart entre réalité et hypothèse nulle et aussi des qualités du terrain, des mesures, et du dispositif expérimental.

Puissance du test de l'effet traitement, en l'absence d'interaction

Sous l'hypothèse d'absence d'interaction, les espérances des carrés moyens concernés par le test (2) (fig. 1), ainsi que leurs degrés de liberté, ne dépendent que du nombre total de parcelles élémentaires par traitement et, bien entendu, des effets principaux (tabl. 2).

TABLEAU 2

Espérances des carrés moyens de l'analyse de variance d'un essai multilocal, en l'absence d'interaction lieu x traitement.

Mean square expectations for the analysis for variance for a multisite experiment in the absence of a site x treatment interaction.

Carré moyen	Espérance	D.d.l.
CM_a	$\sigma^2_E + rq \sum_{i=1}^{P} \frac{a_i^2}{(p-1)}$	(p - 1)
CM_E	σ^2_E	q (p - 1) (r - 1)

Le moins coûteux serait alors de concentrer tout l'essai dans un même lieu, et donc de ne plus faire d'essai multilocal. Toutefois, il est bien clair que si on met en place des essais multilocaux, c'est qu'on ne peut écarter *a priori* l'existence d'une interaction. On ne peut donc pas envisager d'optimiser le dispositif multilocal en négligeant l'interaction.

Puissance du test de détection de l'interaction

Le rapport (CM_{aB}/CM_E) qui permet de tester l'existence d'une interaction traitement x lieu, suit une loi de Fisher-Snedecor à υ_1 et υ_2 degrés de liberté (d.d.l.), multipliée par la constante :

$$1 + r \frac{\sigma^2_{aB}}{\sigma^2_E}$$ (PEARSON et HARTLEY, 1976),

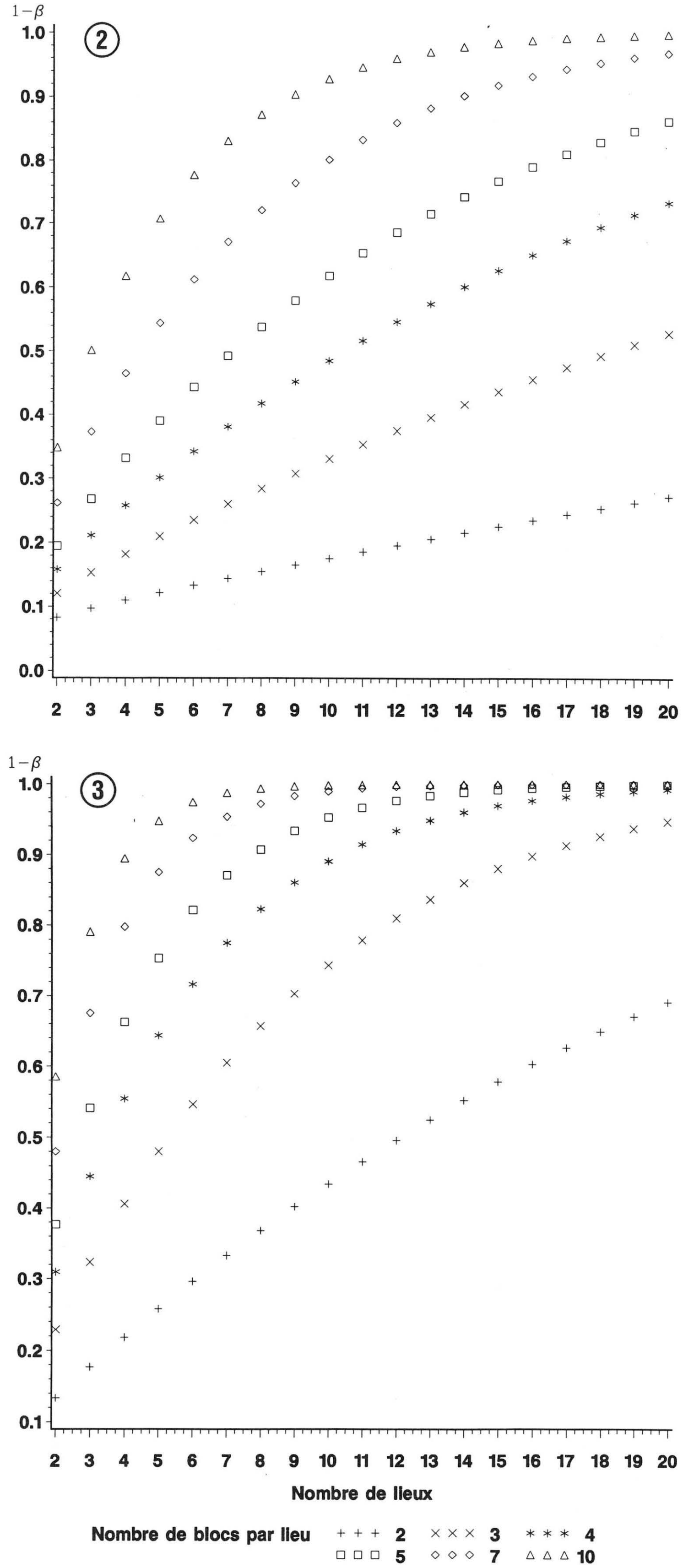
avec $\upsilon_1 = (p-1)(q-1)$
et $\upsilon_2 = q(p-1)(r-1)$

Avec un test au niveau α, l'interaction est détectée quand ce rapport (CM_{aB}/CM_E) dépasse le quantile 1- α de la loi de Fisher à υ_1 et υ_2 d.d.l. On peut donc facilement calculer la puissance du test à partir de la fonction de répartition de cette loi de Fisher. Elle dépendra du niveau α, du nombre d'objets (p), de lieux (q), de répétitions par lieu (r), et du rapport de la variance d'interaction à la variance résiduelle.

* Exemple : en se référant à l'analyse du rendement coton-graine dans des essais génétiques coton en milieu paysan au Togo, on fixe p = 3 traitements comparés et σ_E = 160 kg/ha pour l'écart-type résiduel, ce qui est une valeur courante. Supposons qu'on veuille détecter des interactions qui dépasseraient 150 kg/ha dans 5% des combinaisons lieux x variétés. L'écart-type d'interaction vaudrait 150/1,960 ≈ 75 kg/ha. Le rapport variance d'interaction / variance résiduelle vaut environ $(75/160)^2$ ≈ 0,2. On représente (figures 2 et 3) la puissance du test en fonction de q et de r, et pour deux valeurs du rapport de variances σ^2_{aB}/σ^2_E : 0,2 et 0,5, correspondant à des situations vraisemblables dans les essais multilocaux.

En se reportant à la figure 2, pour 7 lieux, 2 blocs par lieu, et un rapport de variances σ^2_{aB}/σ^2_E = 0,2, la puissance du test vaut environ 14 %. Pour 17 lieux et 2 blocs par lieu, elle n'atteint que 24 %. Le passage à 3 blocs par lieu fait croître la puissance à 26 et 48 % respectivement, ce qui est encore peu.

Dans les essais en milieu contrôlé, avec un écart-type résiduel de 110 kg/ha, σ^2_{aB}/σ^2_E ≈ 0,5, le test de l'interaction atteint une bonne puissance (80 %) si on prévoit 4 lieux et 7 blocs par lieu (fig. 3).



Figures 2 et 3

Puissance du test de l'interaction traitement x lieu, en fonction du nombre de lieux et du nombre de blocs par lieu ; risque $\alpha = 0{,}05$; 3 traitements.

Fig. 2 : variance de l'interaction traitement x lieu : variance résiduelle = 0,2 ;

Fig. 3 : variance de l'interaction traitement x lieu : variance résiduelle = 0,5.

Power of the treatment x site interaction test, as a function of the number of sites and the number of blocks per site; risk $\alpha = 0.05$; 3 treatments.

Fig. 2 : treatment x site interaction variance: residual variance = 0.2;

Fig. 3: treatment x site interaction variance: residual variance = 0.5.

Dans les essais en milieu paysan, il est rarement envisageable de porter à 4 ou 5 le nombre de répétitions par lieu pour atteindre une puissance de 80 %. Compte tenu des contraintes pratiques, le nombre d'essais menés à leur terme ne pourrait être suffisant pour atteindre cet objectif de puissance.

Il n'est donc pas toujours possible de bien tester l'existence d'une interaction avec les moyens dont on dispose.

Aussi, *le résultat négatif du test d'interaction n'est pas toujours probant.* Or, dans l'espérance du carré moyen objets, intervient non seulement l'effet des traitements, mais aussi l'interaction objet x lieu ; le test de l'effet principal par le rapport CM_a/CM_E peut donc être positif à cause d'un effet des traitements, mais aussi à cause d'une interaction qui n'aurait pas été détectée par le test précédent. *Par conséquent, il est prudent de toujours tester l'effet principal à l'aide de $F=CM_d/CM_{aB}$, même si il n'a pas été détecté d'interaction.*

Quand ce test est positif, nous ne pourrons donc affirmer qu'un effet moyen de l'innovation, sans garantir qu'il soit le même dans toutes les situations. *Il convient d'être d'autant plus vigilant sur la représentativité des implantations.*

C'est dans cette perspective de régionalisation de la réponse que nous avons calculé la puissance du test de l'effet principal.

Puissance du test de l'effet principal quand la réponse est régionalisée

Le test de l'effet traitement se fait ici en comparant les carrés moyens CM_a et CM_{aB}, dont les espérances ont été données dans l'introduction.

Le carré moyen CM_a est une somme de variables normales non centrées, puisque les effets a_i ne sont pas nuls. Il suit une loi de $\chi 2$ non centrale, multipliée par la constante σ^2_{aB}.

Le rapport (CM_a/CM_{aB}) suit une loi de Fisher non centrale à υ_1 et υ_2 d.d.l.,

avec $\upsilon_1 = (p-1)$
et $\upsilon_2 = (p-1)(q-1)$,

et de paramètre de non-centralité λ (PEARSON et HARTLEY, 1976)

$$\lambda = r\,q \frac{\Sigma a_i^2}{\sigma^2_E + r.\sigma^2_{aB}}$$

Avec un test au niveau α, l'effet traitement est mis en évidence quand le rapport (CM_a/CM_{aB}) dépasse le quantile 1- α de la loi de Fisher (centrale) à υ_1 et υ_2 d.d.l.

On peut ainsi calculer la puissance de ce test, qui va dépendre du niveau α, de σ^2_E, σ^2_{aB}, Σa_i^2, p, q et r.

En première approximation :

- si σ^2_{aB} est grand devant σ^2_E (forte interaction ou grande précision de chaque résultat élémentaire), alors le terme σ^2_E disparaît devant $r\sigma^2_{aB}$. La puissance ne dépend plus que du nombre d'essais. On aurait alors intérêt à ne faire qu'une répétition par lieu, ce que l'on est amené à éviter pour d'autres raisons (impossibilité de recouper localement les résultats) ;

- si σ^2_{aB} est petit devant σ^2_E, le terme $r\sigma^2_{aB}$ disparaît (car r est faible) devant σ^2_E; l'espérance de CM_a ne dépend plus que du nombre total de parcelles élémentaires (rq). Toutefois, q intervient aussi dans le nombre de degrés de liberté du dénominateur, l'optimum théorique n'est plus la concentration de tout l'essai dans un même lieu, comme c'était le cas pour CM_a/CM_E (2) (fig. 1).

Si on fixe le nombre et les effets des traitements, ainsi que la variance résiduelle, c'est donc la variance d'interaction qui va déterminer le nombre optimal de lieux.

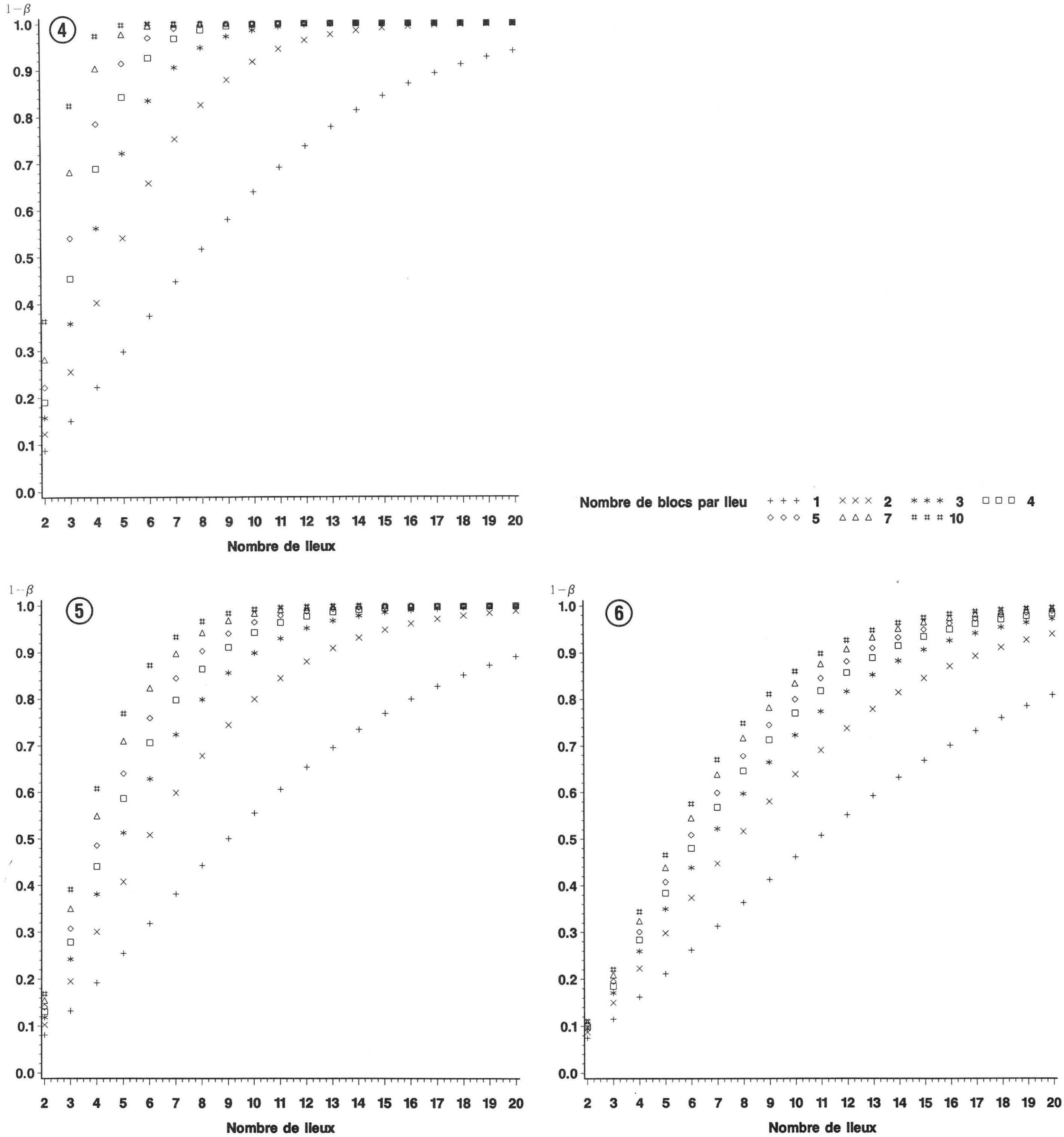
Fixons ces paramètres pour calculer la puissance, en se référant aux mêmes essais génétiques du Togo : p = 3 variétés, σ^2_E = 25 000 kg^2/ha^2, soit $\sigma^2 \approx$ 160 kg/ha. Supposons que les effets des variétés soient respectivement -100, 0, et +100 kg/ha, ce qui est assez considérable pour devoir être détecté. Etudions alors les variations de la puissance en fonction du nombre de lieux et du nombre de blocs par lieu, pour différentes valeurs de la variance d'interaction : 0, 5 000 et 12 500 kg^2/ha^2, soit des écart-types d'interaction d'environ 0, 70 et 110 kg/ha (fig. 4 à 6).

Ces graphes montrent des caractéristiques assez simples :
- la puissance décroît quand la variance d'interaction croît ;

- la puissance croît quasi-linéairement en fonction du nombre de lieux, jusqu'à environ 70 % ;

- la puissance augmente sensiblement de 1 répétition par lieu (blocs dispersés) à 2, puis 3 répétitions par lieu. Le gain est d'autant moins fort que la variance d'interaction est élevée, mais il demeure substantiel même pour des valeurs élevées de la variance d'interaction.

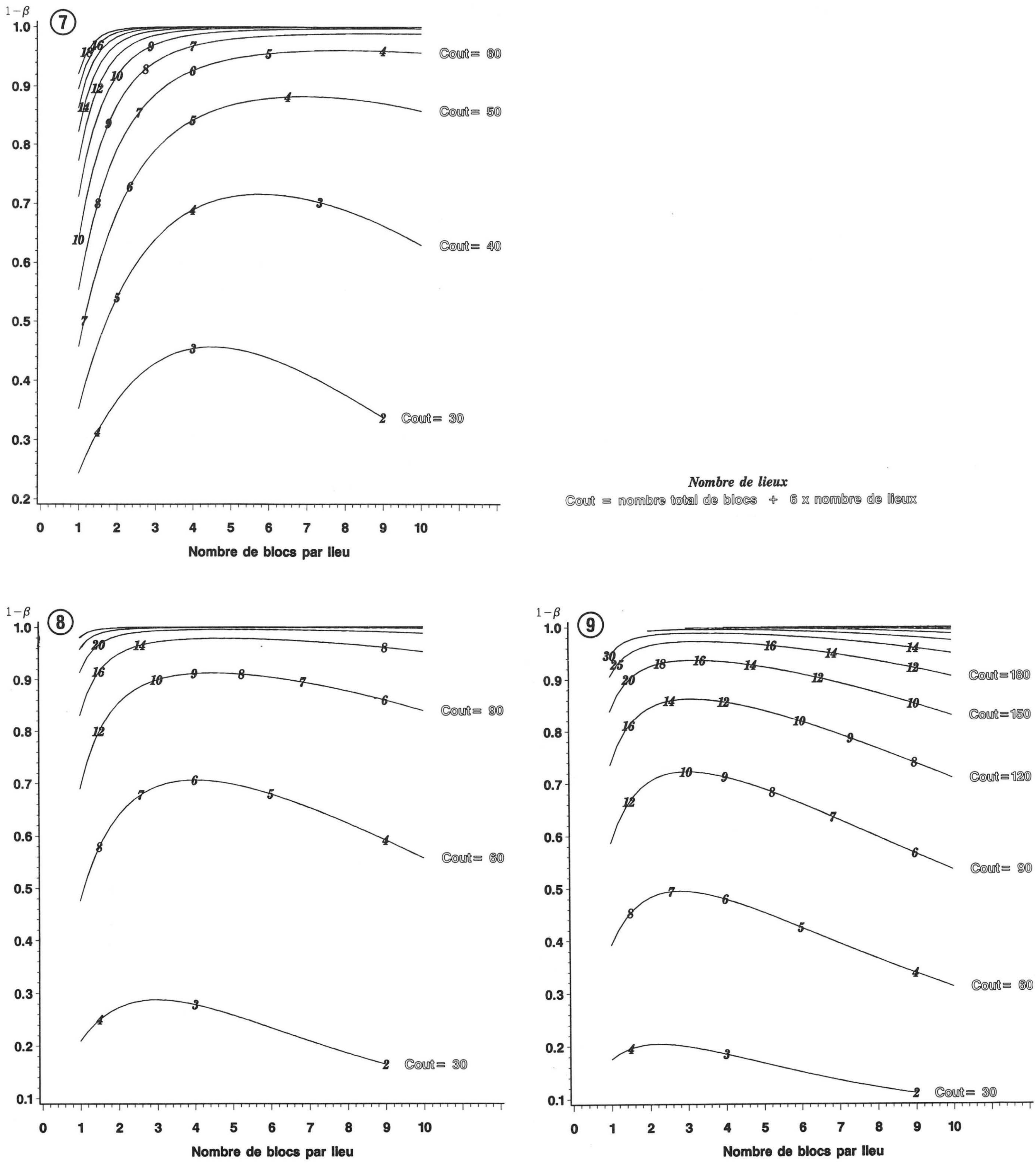
Compte tenu d'un coût marginal assez faible d'une répétition supplémentaire, par rapport à celui d'un lieu supplémentaire dans la situation étudiée, le passage de 1 à 2 puis 3 répétitions par lieu est justifié, pour autant que les conditions expérimentales le permettent.



Figures 4, 5 et 6

Puissance du test de l'effet traitement, en fonction du nombre de lieux et du nombre de blocs par lieu ; risque $\alpha = 5\%$; 3 traitements comparés d'effets -100, 0 et +100 ; variance résiduelle = 25 000.
Fig. 4 : variance d'interaction traitement x lieu = 0 ;
Fig. 5 : variance d'interaction traitement x lieu = 5 000 ;
Fig. 6 : variance d'interaction traitement x lieu = 12 500.

Power of the treatment effect test, as a function of the number of sites and the number of blocks per site ; risk $\alpha = 5\%$; 3 treatments effects -100, 0 and + 100 ; residual variance = 25,000.
Fig. 4: treatment x site interaction variance = 0;
Fig. 5: treatment x site interaction variance = 5,000;
Fig. 6: treatment x site interaction variance = 12,500.



Figures 7, 8 et 9

Puissance du test de l'effet traitement, en fonction du coût et du nombre de blocs par lieu ; risque $\alpha = 5\%$; 3 traitements comparés d'effets -100, 0 et +100 ; variance résiduelle = 25 000.

Fig. 7 : variance d'interaction traitement x lieu = 0 ;

Fig. 8 : variance d'interaction traitement x lieu = 5000 ;

Fig. 9 : variance d'interaction traitement x lieu = 12 500.

Power of the treatment effect test, as a function of cost and of the number of blocks per site; risk $\alpha = 5\%$; 3 treatments, effects -100, 0 and +100, residual variance = 25,000.

Fig. 7: treatment x site interaction variance = 0;

Fig. 8: treatment x site interaction variance = 5,000;

Fig. 9: treatment x site interaction variance = 12,500.

Exemple d'optimisation économique pour le test de l'effet principal

Il reste à moduler cette affirmation, en fonction du coût relatif des lieux, des blocs et de l'importance de l'interaction, pour trouver, en fonction des moyens mis en oeuvre, la meilleure combinaison (nombre de lieux, nombre de blocs par lieu), en supposant que ces nombres ne soient pas limités par des contraintes autres que budgétaires.

Il nous faut pour cela faire une hypothèse sur les coûts : nous considérerons ici, à titre d'exemple, que la mise en place d'un lieu supplémentaire coûte 6 fois plus cher que celle d'un bloc supplémentaire dans un lieu quelconque.

Le coût global de l'expérimentation est alors proportionnel à (6+r)q.

La meilleure valorisation des moyens mis en jeu est atteinte quand, pour un coût donné, on atteint la meilleure puissance. Le problème n'est pas très simple à résoudre, même numériquement, car les nombres de lieux et de blocs par lieu sont forcément entiers : on ne peut donc envisager, en se fixant un coût, tous les nombres de lieux possibles.

On détermine l'optimum graphiquement, comme si les nombres de lieux et de blocs par lieu pouvaient varier continûment, puis on arrondit la solution trouvée. Cette option est possible car la puissance est une fonction strictement croissante de chacun de ces paramètres.

Les figures 7 à 9 sont des courbes de puissance à coût constant, fonction du nombre de répétitions par lieu. Pour chaque courbe, quand le nombre de blocs par lieu croît, le nombre de lieux décroît, puisque le coût est constant. Les nombres de lieux figurent en italique sur les courbes. Ils sont donnés simplement à titre indicatif, car ils ne correspondent pas toujours à un nombre entier de blocs par lieu (cette contrainte ne permettrait pas de tracer des courbes à coût constant et de déterminer par ce moyen les optima).

Pour utiliser ces courbes, on détermine le budget dont on dispose et on choisit la courbe correspondant au coût le plus proche. On déduit alors le nombre optimal de répétitions par lieu, qu'il faut arrondir à la plus proche valeur entière. On peut trouver aussi le nombre approximatif de lieux. Le nombre de lieux définitif doit être recalculé en fonction du nombre (entier) de répétitions par lieu pour ne pas dépasser le budget initial.

Comme on peut s'y attendre, le nombre optimal de blocs par lieu s'abaisse quand la variance d'interaction s'accroit : en absence d'interaction, il dépasse 6 blocs par lieu (fig. 7), pour atteindre 3 blocs par lieu quand l'interaction devient forte (σ_{aB} = 110 kg/ha, fig. 9).

Dans le même temps, l'optimum de puissance devient moins marqué. Quand la variance d'interaction est faible, l'optimisation à coût constant peut donner des gains de puissance de 35 % ; quand l'interaction est forte, les gains potentiels sont beaucoup plus modestes (10 à 15 %), ceci étant aussi à relier à une baisse générale de la puissance.

Par rapport aux dispositifs habituellement mis en place à l'IRCT, pour l'étude du rendement coton-graine, quelle que soit la variance d'interaction envisagée ici, quand la puissance potentielle correspond à un objectif raisonnable (70 %), il est toujours payant d'augmenter le nombre de blocs par lieu, au moins jusqu'à 3. Les blocs dispersés, en particulier, représentent, pour un coût donné, une perte de puissance importante par rapport aux essais à 2 ou 3 répétitions par lieu.

Discussion

L'utilisation pratique de ces résultats suppose :

- de connaître la variance résiduelle ;

- d'admettre que l'interaction est répartie suivant une loi normale, et de connaître aussi sa variance.

D'autre part, on n'a envisagé ici qu'une seule variable à étudier, ce qui n'est pas le cas général.

Le premier point n'amène que la difficulté de réunir les résultats d'essais déjà réalisés dans les mêmes conditions pour estimer la variance résiduelle ou, à défaut, d'en donner une valeur raisonnable.

Le second est sans doute le plus délicat :

- l'hypothèse de normalité est vérifiée quand plusieurs causes indépendantes d'interaction viennent ajouter des effets de même ordre de grandeur. Elle ne l'est pas en cas de séparation en zones très contrastées (types de sols, par exemple), ou encore en cas de gradient (de pluviométrie, par exemple). La régionalisation de la réponse n'intervient alors qu'après modélisation convenable de l'interaction, pour en éliminer les composantes majeures (DENIS, VINCOURT, 1982) ;

- la variance d'interaction peut être fixée à un seuil jusqu'auquel les disparités entre situations locales sont tolérables. La taille du réseau est alors calculée pour détecter les interactions supérieures à ce seuil, et pour que les effets principaux soient mis en évidence quand l'interaction est inférieure à ce même seuil.

Le choix de celui-ci est déterminant pour la forme d'un réseau : tolérer une forte interaction conduit à mettre en place de nombreux essais à peu de répétitions, au risque d'une dépense injustifiée si l'interaction est en réalité faible. Au contraire, miser sur une faible interaction amène à concentrer les essais dans quelques lieux, au risque de rater son objectif de puissance si l'interaction est plus élevée que prévu.

Enfin, l'optimisation d'un réseau pour l'étude de plusieurs variables peut être décomposée de la façon suivante :

- pour chacune des variables, on se fixe un objectif de puissance, auquel correspond un ensemble de situations acceptables (nombre q de lieux, nombre r de blocs par lieu) ;

- toutes les fonctions de puissance étant croissantes en fonction de q et de r, ces ensembles admettent forcément une intersection non vide ;

- dans cette intersection, on recherche alors la combinaison de coût minimum qui donne le nombre d'essais à mettre en place et le nombre de répétitions par essai. Ces nombres étant entiers, la mise en oeuvre informatique des opérations d'intersection et de recherche du minimum est assez simple ;

- pour chacune des variables où l'objectif de puissance est dépassé, on peut ensuite diminuer les coûts de prise d'échantillons en éliminant des répétitions. Si le coût de la mesure est déterminant, on peut regrouper dans un même échantillon plusieurs prélèvements d'un même lieu.

Conclusion

La décision de diffuser une innovation se fait après un essai multilocal que l'on peut considérer comme un sondage.

Par ce sondage dans une région supposée homogène, on vérifie l'avantage comparatif global de l'innovation proposée. On peut admettre que cet avantage ne soit pas constant dans toute la région, en tolérant une interaction objet x lieu, jusqu'à un certain niveau.

Aussi bien pour la mise en évidence de cette interaction que pour celle de l'effet global, on peut, à partir d'hypothèses simples, déterminer la dimension d'un réseau d'essais.

Références bibliographiques

DAGNELIE P., 1973.- Les tests d'hypothèses. *In* : Théorie et méthodes statistiques. *Vander, Bruxelles (B)*, 2e éd., vol. 1, 327-353.

DENIS J.B., VINCOURT P., 1982.- Panorama des méthodes statistiques d'analyse des interactions génotype x milieu. *Agronomie*, 2, 3, 219-230.

PEARSON E.S., HARTLEY H.O., 1976.- Biometrika tables for statisticians. *Biometrika Trust, London (GB)*, vol. 2, 385 p.

PHILIPPEAU G., 1984.- Puissance d'une expérience. Nombre de répétitions nécessaires pour comparer deux ou plusieurs traitements. *ITCF, service des études statistiques, 8 av. du Président Wilson, Paris*, 20 p.

— — — — — — — —

Determining the size of multisite experiments

E. Gozé

Abstract

The size of an experiment is calculated using a power target. However, the power of a multisite experiment is rarely known, either for detecting treatment-site interactions or the main effects. A reminder of the aim of multisite experiments, the model and the analysis methods is followed by the formulae for calculating the power of a series of complete-block experiments. The model uses random site effects and fixed treatment effects. In the case of trials with 3 treatments, the power curves for the different tests are drawn up and commented on. An example of economic optimization is then given. To detect the main effects on seed cotton yields, a substantial increase in power can be achieved by replacing scattered blocks with designs comprising 2 or 3 replicates per site.

KEY WORDS: cotton, multisite experiments, smallholder experiments, power, analysis of variance, regionalization, interaction, optimization.

Introduction

Aim

The aim of multisite experiments is to test one or more innovations (varieties, crop techniques, phytosanitary measures) under various environmental conditions (climate, soil, smallholders). The innovation may be accepted or rejected after such an experiment, depending on decisive qualitative criteria: growth habit of a variety, sensitivity to deficiencies, toxicity of a phytosanitary product, degree of acceptability to farmers, etc. The choice of one or other of the treatments tested can also be governed by quantitative criteria, such as measurement of infestations, fibre quality and yields, as far as cotton is concerned. We shall be looking at these quantitative criteria here.

The innovation, either alone or in comparison with others, is generally compared with a control. The aim is to determine whether it represents an economic gain over the whole of the zone covered by the multisite experiment (which does not mean that the innovation will have a positive effect for all the farmers in the zone). If it is to be adopted, the innovation has to:

- represent a gain in every possible situation (case A);

- or provide sufficient gains in most situations for possible losses to be more than compensated for over the zone as a whole (case B). The aim is not to make use of local differences.

Model

In order to draw conclusions for the whole of the zone studied, experiment has to be carried out at a sample of sites representative of the possible situations, for which a quantitative variable - y - is measured. In the absence of existing information better suited to the phenomenon studied, the true situation is represented by a fairly general linear model:

$$y_{ijk} = m + a_i + B_j + (aB)_{ij} + c_{jk} + E_{ijk}$$

where

m	is the mean effect		
a_i	is the effect of treatment i	(fixed),	i=1..p treatments
B_j	is the effect of site j	(random),	j=1..q sites
$(aB)_{ij}$	is the treatment x site interaction	(random)	
c_{jk}	is the effect of block k at site j	(fixed)	k=1..r blocks/site
E_{ijk}	is the residual error	(random)	

It is assumed that the residual errors are normally distributed about a zero expectation, that they are independent, that their variance is the same everywhere, and that the same goes for the terms of interaction. Moreover, errors and interactions are assumed to be independant of one another.

Analysis of variance

The aims described above boil down to the detection of the main effects and interactions, which is done using statistical tests. The analysis of variance makes it possible to calculate mean squares, whose expectation and degrees of freedom are given in table 1. These mean squares are quoted MS in the text and CM in the tables and figures.

Just as σ^2_E is the variance of errors E_{ijk}, σ^2_{aB} is the variance of the terms of interaction $(aB)_{ij}$, and these terms are assumed to be normally distributed.

If this variance is nil, the effects of the treatments and the sites are strictly additive: the differences between the objects are the same at each site, apart from residual errors.

If the variance of the interaction is 100, 50% on average of these $(aB)_{ij}$ exceed $0.675 \times \sqrt{100} = 6.75$ as an absolute value and 5% exceed $1.960 \times \sqrt{100} = 19.6$.

It is worth noting that the mean square for the interaction is not equal to the variance of the interaction. To estimate this variance based on the results of an experiment, the unbiased estimator is used:

$$\hat{\sigma}^2_{aB} = \frac{MS_{aB} - MS_E}{r}$$

Tests

As shown in figure 1, if no interaction is detected, it is concluded that there is none. Any gains made as a result of the innovation will therefore assumed to be the same for all the sites. The treatment effect test is carried out using the ratio MS_a/MS_E.

However, if there is an interaction, a mean effect can be detected over the whole of the zone covered by the experiment, using the ratio MS_a/MS_{aB}: this is known as regionalizing the response.

The problem we propose to look at here is how to determine the number of trials and blocks required to maximize the chances of detecting the interaction and the main effects, using as few resources as possible.

In a statistical test, the probability of rejecting the null hypothesis (in this case of detecting an effect) is known as the power (see PHILIPPEAU, 1984 or DAGNELIE, 1973). It depends on the extent of the difference between the true situation and the null hypothesis, and also on the quality of the land, measurements, and experimental design.